A Theory of Parallelism and the Case of VP Ellipsis

Jerry R. Hobbs and Andrew Kehler
Artificial Intelligence Center
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025
{hobbs,kehler}@ai.sri.com

Abstract

We provide a general account of parallelism in discourse, and apply it to the special case of resolving possible readings for instances of VP ellipsis. We show how several problematic examples are accounted for in a natural and straightforward fashion. The generality of the approach makes it directly applicable to a variety of other types of ellipsis and reference.

1 The Problem of VP Ellipsis

VP ellipsis has received a great deal of attention in theoretical and computational linguistics (Asher, 1993; Crouch, 1995; Dalrymple, Shieber, and Pereira, 1991; Fiengo and May, 1994; Gawron and Peters, 1990; Hardt, 1992; Kehler, 1993; Lappin and McCord, 1990; Prüst, 1992; Sag, 1976; Webber, 1978; Williams, 1977, inter alia). The area is a tangled thicket of examples in which readings are mysteriously missing and small changes reverse judgments. It is a prime example of a phenomenon at the boundary between syntax and pragmatics.

VP ellipsis is exemplified in sentence (1).

(1) John revised his paper before the teacher did.

This sentence has two readings, one in which the teacher revised John's paper (the *strict* reading), and one in which the teacher revised his own paper (the *sloppy* reading). Obtaining an adequate account of strict/sloppy ambiguities has been a major focus of VP ellipsis research. This is challenging because not all examples are as simple as sentence (1). In fact, sentence (1) is the first main clause of one of the more problematic cases in the literature:

(2) John revised his paper before the teacher did, and Bill did too.

Whereas one might expect there to be as many as six readings for this sentence, Dalrymple et al. (1991, henceforth DSP) note that it has only five readings; the reading is absent in which

(3) John revised John's paper before the teacher revised John's paper, and Bill revised John's paper before the teacher revised Bill's paper.

Previous analyses have either generated too few or too many readings, or have required an appeal to additional processes or constraints external to the actual resolution process itself.

Examples like (2) test the adequacy of an analysis at a fine-grained level of detail. Other examples test the generality of an analysis, in terms of its ability to account for phenomena similar to VP ellipsis and to interact with other interpretation processes that may come into play. For instance, strict/sloppy ambiguities are not restricted to VP ellipsis, but are common to a wide range of constructions that rely on parallelism between two eventualities, some of which are listed in Table 1. Given the ubiquity of strict/sloppy ambiguities, one would expect these to be a by-product of general discourse resolution mechanisms and not mechanisms specific to VP ellipsis. Any account applying only to the latter would miss an important generalization.

In this paper, we give an account of resolution rooted in a general computational theory of parallelism. We demonstrate the depth of our approach by showing that unlike previous approaches, the algorithm generates the correct five readings for example (2) without appeal to additional mechanisms or constraints. We also discuss how other 'missing readings' cases are accounted for. We show the generality of the approach by demonstrating its handling of several other examples that prove problematic for past approaches, including a *source-of-ellipsis paradox*, so-called *extended parallelism* cases, and *sloppy readings with events* cases. Of the phe-

Phenomenon	Example
'Do It' Anaphora	John revised his paper before Bill did it.
'Do So' Anaphora	John revised his paper and Bill did so too.
Stripping	John revised his paper, and Bill too.
Comparative Deletion	John revised his paper more quickly than Bill.
'Same As' Reference	John revised his paper, and Bill did the same.
	John revised his paper, and the teacher followed suit.
'Me Too' Phenomena	A: John revised his paper. B: Me too./Ditto.
'One' Anaphora	John revised a paper of his, and Bill revised one too.
Lazy Pronouns	The student who revised his paper did better than the student who handed it in as is.
Anaphoric Deaccenting	John said he called his teacher an idiot, and Bill *said he insulted his teacher* too.
Focus Phenomena	Only John revised his paper.

Table 1: Phenomena Giving Rise to Sloppy Interpretations

nomena in Table 1, we briefly discuss the algorithm's handling of *lazy pronoun* cases.

2 A Theory of Parallelism

The Theory A clause conveys a property or eventuality, or describes a situation, or expresses a proposition. We use the term "property" to cover all of these cases. A property consists of a predicate applied to a number of arguments. We make use of a duality between properties having a number of arguments, and arguments having a number of properties. Parallelism is characterized in terms of a co-recursion in which the similarity of properties is defined in terms of the similarity of arguments, and the similarity of arguments is defined in terms of the similarity of properties.[1]

Two fragments of discourse stand in a parallel relation if they describe similar properties. Two properties are similar if two corresponding properties can be inferred from them in which the predicates are the same and the corresponding pairs of arguments are either coreferential or similar.

$$\begin{array}{l} Similar[p'_1(e_1, x_1, \ldots, z_1), p'_2(e_2, x_2, \ldots, z_2)]: \\ \quad p'_1(e_1, x_1, \ldots, z_1) \supset p'(e_1, x_1, \ldots, z_1) \text{ and} \\ \quad p'_2(e_2, x_2, \ldots, z_2) \supset p'(e_2, x_2, \ldots, z_2) \text{, where} \\ \qquad Coref(x_1, \ldots, x_2, \ldots) \text{ or } Similar[x_1, x_2], \\ \qquad \ldots \\ \qquad Coref(z_1, \ldots, z_2, \ldots) \text{ or } Similar[z_1, z_2] \end{array}$$

Two arguments are similar if their other, "inferentially independent" properties are similar.

$$\begin{array}{l} Similar[x_1, x_2]: \\ \quad Similar[p'_1(\ldots, x_1, \ldots), p'_2(\ldots, x_2, \ldots)], \\ \quad \ldots \\ \quad Similar[q'_1(\ldots, x_1, \ldots), q'_2(\ldots, x_2, \ldots)] \end{array}$$

[1]This account is a elaboration of treatments of parallelism by Hobbs (1979; 1985) and Kehler (1995).

The constructed mapping between pairs of arguments must be preserved and remain one-to-one.

There are three ways the recursion can bottom out. We can run out of new arguments in properties. We can run out of new, inferentially independent properties of arguments. And we can "bail out" of proving similarity by proving or assuming coreference between the two entities.

Two properties are *inferentially independent* if neither can be derived from the other. Given a knowledge base K representing the mutual knowledge of the participants in the discourse, properties P_1 and P_2 are inferentially independent if neither $K, P_1 \vdash P_2$ nor $K, P_2 \vdash P_1$. This rules out the case in which, for example, the fact that John and Bill are both persons would be used to establish their similarity when the fact that they are both men has already been used. Inferential independence is generally undecidable, but in practice this is not a problem. In discourse interpretation, all we usually know about an entity is the small set of properties presented explicitly in the text itself. We may take these to be inferentially independent and look for no further properties, once properties inferrable from these have been used in establishing the parallelism.

Similarity is a matter of degree. The more corresponding pairs of inferentially independent properties that are found, and the more contextually salient those properties are, the stronger the similarity. In a system which assigns different costs to proofs (e.g., Hobbs et al. (1993)), the more costly the proofs required to establish similarity are, the less similar the properties or arguments should seem. Interpretations should seek to maximize similarity.

This account of parallelism is semantic in the sense that it depends on the content of the discourse rather than directly on its form. But syntax plays an implicit role. When seeking to establish the parallelism between two clauses, we must begin with the "top-level" properties; this is generally determined by the syntactic structure of the clause. Then the co-recursion through the arguments and properties normally mirrors the syntactic structure of the sentence. However, features of syntax that are not manifested in logical form are not taken into account.

An Example To illustrate that the theory has applicability well beyond the problem of VP ellipsis, we present an example of semantic parallelism in discourse. It comes from an elementary physics textbook, and is worked out in essentially the same manner in Hobbs (1979).

(4) A ladder weighs 100 lb with its center of gravity 10 ft from the foot, and a 150 lb man is 10 ft from the top.

We will assume "the foot" has been identified as the foot of the ladder. Because it is a physics problem, we must reduce the two clauses to statements about forces acting on objects with magnitudes in a direction at a point in the object:

$$force(w_1, L, d_1, x_1); force(w_2, y, d_2, x_2)$$

In the second clause we do not know that the man is standing on the ladder—he could be on the roof—and we do not know what "the top" is the top of. These facts fall out of recognizing the parallelism.

The procedure for establishing parallelism is illustrated in Figure 1, in which parallel elements are placed on the same line. The *force* predicates are the same so there is no need to infer further properties. The first pair of arguments, w_1 and w_2 are similar in that both are weights. To make the second pair of arguments similar, we can assume they are coreferential; as a by-product, this tells us that the object the man's weight is acting on is the ladder, and hence that the man is on the ladder. The third pair of arguments are both downward directions. The final pair of arguments, x_1 and x_2, are similar if their properties $distance(x_1, f, 20ft)$ and $distance(x_2, t, 10ft)$ are similar. These will be similar if their previously unmatched pair of arguments f and t are similar. This holds if their properties $foot(f, L)$ and $top(t, z)$ are similar. We infer $end(f, L)$ and $end(t, z)$, since feet and tops are ends. Finally, we have to show L and z are similar. We can do this by assuming they are coreferential. This, as a by-product, tells us that the top is the top of the ladder.

The use of inferences, such as "a foot is an end", means that this theory is parametric on a knowledge base. Different sets of beliefs can yield different bases for parallelism and indeed different judgments about whether parallelism occurs at all.

A crucial piece of our treatment of VP-ellipsis is the explicit representation of coreference relations, denoted with the predicate *Coref*. We could use equalities such as $y = L$, or since equals can be replaced by equals, simply replace y with L. However, doing this would lose the distinction between y and L under their corresponding descriptions.

Consequently, we introduce the relation $Coref(y, e_2, x, e_1)$ to express this coreferentiality. This relation says that y under the description associated with e_2 is coreferential with x under the description associated with e_1. From this we can infer $y = x$ but not $e_2 = e_1$, and the coreferentiality cannot be washed out in substitution. A constraint on the arguments of *Coref* is that e_1 and e_2 be properties of x and y respectively.

The phenomenon of parallelism pervades discourse. In addition to straightforward examples of parallelism like the above, there are also contrasts, exemplifications, and generalizations, which are defined in a similar manner. The interpretation of a number of syntactic constructions depends on recognizing parallelism, including those cited in Table 1. In brief, our theory of parallelism is not something we have introduced merely for the purpose of handling VP ellipsis; it is needed for a wide range of sentential and discourse phenomena.

Other Approaches Based on Parallelism Our aim in this paper is to present the theory of parallelism at an abstract enough level that it can be embedded in any sufficiently powerful framework. By "sufficiently powerful" we mean that there must be a formalization of the notion of inference, strength of inference, and inferential independence, and there must be a reasonable knowledge base. In Hobbs and Kehler (forthcoming), we show how our approach can be realized within the "Interpretation as Abduction" framework (Hobbs et al., 1993).

There are at least two other treatments in which VP ellipsis is resolved through a more general system of determining discourse parallelism, namely, those of Prüst (1992) and Asher (1993).

Prüst (1992) gives an account of parallelism developed within the context of the Linguistic Discourse Model theory (Scha and Polanyi, 1988). Parallelism is computed by determining the "Most Specific Common Denominator" of a set of representations, which results from unifying the unifiable aspects of those representations and generalizing over the others. VP ellipsis is resolved as a side effect of this unification. The representations assumed, called *syntac-*

$$
\begin{array}{l}
force(w_1, L, d_1, x_1) \\
\quad w_1 : lb(w_1, 100) \\
\quad\quad L : ladder(L) \\
\quad\quad d_1 : Down(d_1) \\
\quad\quad x_1 : distance(x_1, f, 20ft) \\
\quad\quad\quad f : foot(f, L) \Rightarrow end(f, L) \\
\quad\quad\quad\quad L :
\end{array}
\qquad\qquad
\begin{array}{l}
force(w_2, y, d_2, x_2) \\
\quad w_2 : lb(w_2, 150) \\
\quad\quad y :\Rightarrow Coref(y, ..., L, ...) \\
\quad\quad d_2 : Down(d_2) \\
\quad\quad x_2 : distance(x_2, t, 10ft) \\
\quad\quad\quad t : top(t, z) \Rightarrow end(t, z) \\
\quad\quad\quad\quad z :\Rightarrow Coref(z, ..., L, ...)
\end{array}
$$

Figure 1: Example of Parallelism Establishment

tic/semantic structures, incorporate both syntactic and semantic information about an utterance. One weakness of this approach is that it appears overly restrictive in the syntactic similarity that it requires.

Asher (1993) also provides an analysis of VP ellipsis in the context of a theory of discourse structure and coherence, using an extension of Discourse Representation Theory. The resolution of VP ellipsis is driven by a need to maximize parallelism (or in some cases, contrast) that is very much in the spirit of what we present.

Detailed comparisons with our approach are given with the examples below. In general, however, in neither of these approaches has enough attention been paid to other interacting phenomena to explain the facts at the level of detail that we do.

3 VP Ellipsis: A Simple Case

We first illustrate our approach on the simple case of VP ellipsis in sentence (1). The representation for the antecedent clause in our "logical form"[2] appears on the left-hand side of Figure 2. Note that a *Coref* relation links x_1, the variable corresponding to "he" (eventuality e_{13}), to its antecedent j, the entity described by "John" (eventuality e_{11}).

From the second clause we know there is an elided eventuality e_{22} of unknown type P, the logical subject of which is the teacher t.

$$
\begin{array}{l}
P(e_{22}, t) \\
\quad t : teacher'(e_{21}, t)
\end{array}
$$

Because of the ellipsis, e_{22} must stand in a parallel relation to some previous eventuality; here the only candidate is John's revising his paper (e_{12}). To establish $Similar(e_{12}, e_{22})$,[3] we need to show that their corresponding arguments are similar. John j and the teacher t are similar by virtue of being persons. The corresponding objects p_1 and p_2 are similar if we take p_2 to be a paper and to have a *Poss* property similar to that of p_1. The latter is true if corresponding to the possessor x_1, there is an x_2 that is similar to x_1.

In constructing the similarity between x_2 and x_1, we can either take them to be coreferential (case *a) or prove them to be similar by having similar properties, including having similar dependencies established by *Coref* (case *b). In the former case, x_2 is coreferential with x_1 which is coreferential with John j, giving us the strict reading. In the latter case, we must preserve the previously-constructed mapping between John j (on which x_1 is dependent) and the teacher t; thus x_2 is similar to x_1 if taken to be coreferential with t, giving us the sloppy reading.[4]

4 A Missing Readings Paradox

Sentence (1) is the antecedent clause for example (2), one of the more problematic examples in the literature. Theoretically, this example could have as many as six readings, paraphrased as follows:

(5) John revised John's paper before the teacher revised John's paper, and Bill revised John's/Bill's paper before the teacher revised John's/Bill's paper.

(6) John revised John's paper before the teacher revised the teacher's paper, and Bill revised John's/Bill's paper before the teacher revised the teacher's paper.

[2]The normally controversial term "logical form" is used loosely here, simply to capture the information that the hearer must bear in mind, at least implicitly, in interpreting texts such as sentence (1).

[3] We cannot establish coreference between the events because their agents are distinct. In other cases, however, the process can bail out immediately in event coreference; consider the sentence "John revised his paper, smoking incessantly as he did." A *Coref* link is established between the elided and antecedent events in the same way as for pronouns. This symmetry accounts for another problematic case, discussed in Section 6.

[4]It is also possible to "bail out" in coreference between the papers p_1 and p_2; here we would get the strict reading again. However, consider if the example had said "a paper of his" rather than "his paper". The resulting sentence has two strict readings, one in which both revised the same paper of John's (generated by assuming coreference between the papers), and one in which each revised a (possibly) different paper of John's (generated by assuming coreference between the pronouns).

$$
\begin{array}{ll}
\begin{array}{l}
before'(e_{12}, e_{22}) \\
\quad revise'(e_{12}, j, p_1) \\
\quad\quad j : John'(e_{11}, j) \\
\quad\quad p_1 : paper'(e_{15}, p_1) \\
\quad\quad\quad Poss'(e_{14}, x_1, p_1) \\
\quad\quad\quad\quad x_1 : he'(e_{13}, x_1) \\
\quad\quad\quad\quad\quad Coref(x_1, e_{13}, j, e_{11})
\end{array}
&
\begin{array}{l}
\\
revise'(e_{22}, t, p_2) \\
\quad t : teacher'(e_{21}, t) \\
\quad p_2 : paper'(e_{25}, p_2) \\
\quad\quad Poss'(e_{24}, x_2, p_2) \\
\quad\quad\quad x_2 : he'(e_{23}, x_2) \\
\quad\quad\quad\quad [Coref(x_2, e_{23}, x_1, e_{13})\ (\text{*a})] \\
\quad\quad\quad\quad [Coref(x_2, e_{23}, t, e_{21})\ (\text{*b})]
\end{array}
\end{array}
$$

Figure 2: Representations for Simple Case

We follow DSP in claiming that this example has five readings, in which the JJJB reading shown in (3) is missing.[5] DSP, who use this case as a benchmark for theories of VP ellipsis, note that the methods of Sag (1976) and Williams (1977) can be seen to derive two readings, namely JJJJ and JTBT. An analysis proposed by Gawron and Peters (1990), who first introduced this example, generates three readings (adding JJBB to the above two), as does the analysis of Fiengo and May (1994). A method that Gawron and Peters attribute to Hans Kamp generates either four readings, including the above three and JTJT, or all six readings. DSP's analysis strictly speaking generates all six readings; however, they appeal to anaphor/antecedent linking relationships to eliminate the JJJB reading. However, these linking relationships are not a by-product of the resolution process itself, but must be generated separately. Our approach derives exactly the correct five readings.[6]

The antecedent clause is represented in Figure 2, and the expansion of the final VP ellipsis is shown in Figure 3. In proving similarity, each pronoun can be taken to be coreferential with its parallel element (cases *a, *c and *e), or proven similar to it (cases *b, *d, *f and *g). If choice *a is taken in the second clause, then the "similarity" choice in the fourth clause must be *f; if *b, then *g. If *a and *c are chosen, the JJJJ reading results. If *a, *d, and *e are chosen, the JJBJ reading results. If *a, *d, and *f are chosen, the JJBB reading results. If *b and *c are chosen, the JTJT reading results. If *b and *d are chosen, the JTBT reading results. Thus taking all possible choices gives us all acceptable readings.

Now consider what it would take to obtain the *JJJB reading. The variable x_3 would have to be coreferential with John and x_4 with Bill. The former requirement forces us to pick case *c. But then case *e makes x_4 coreferential with either John or the teacher (depending on how the first ellipsis was resolved). Case *f makes x_4 coreferential with John, and case *g makes it coreferential with the teacher. There is no way to get x_4 coreferential with Bill once we have set x_3 to something other than Bill.

Neither Prüst (1992) nor Asher (1993) discuss this example. In extrapolating from the analyses Prüst gives, we find that his analysis generates only two of the five readings. Briefly, if the first ellipsis is resolved to the strict reading, then the JJJJ reading is possible. If the first ellipsis is resolved to the sloppy reading, then only the JTBT reading is possible. Asher's account, extrapolating from an example he discusses (p. 371), may generate as many as six readings, including the missing reading. This reading results from the manner in which the strict reading for the first ellipsis is generated—the final clause pronoun is resolved with the entity specified by the subject of the antecedent clause, whereas our algorithm creates a dependency between the pronoun and its parallel element in the antecedent clause. Our mechanism is more natural because of the alignment of parallel elements between clauses when establishing parallelism, and it is this property which results in the underivability of the missing reading.

5 A Source-of-Ellipsis Paradox

DSP identify two kinds of analysis in the VP ellipsis literature. In *identity-of-relations* analyses (Sag, 1976; Williams, 1977; Gawron and Peters, 1990; Fiengo and May, 1994, inter alia) strict/sloppy readings arise from an ambiguity in the antecedent VP derivation. The ambiguity in the ellipsis results from copying each possibility. In *non-identity* approaches (Dalrymple, Shieber, and Pereira, 1991; Kehler, 1993; Crouch, 1995, inter alia) strict/sloppy readings result from a choice point within the resolution algorithm. Our approach falls into this class.

Non-identity approaches are supported by examples such as (7), which has reading (8).

[5] Each reading for this example contains four descriptions of papers that were revised. We use the notation JJJB to represent the reading in which the first three papers are John's and fourth is Bill's, corresponding to reading (3). Other uses of such notation should be understood analogously.

[6] The approach presented in Kehler (1993) also derives the correct five readings, however, our method has advantages in its being more general and better motivated.

$$
\begin{array}{l}
before(e_{32}, e_{42}) \\
\quad revise'(e_{32}, b, p_3) \\
\quad\quad b : Bill'(e_{31}, b) \\
\quad\quad p_3 : paper'(e_{35}, p_3) \\
\quad\quad\quad Poss'(e_{34}, x_3, p_3) \\
\quad\quad\quad\quad x_3 : he'(e_{33}, x_3) \\
\quad\quad\quad\quad\quad [(\text{*c})\ Coref(x_3, e_{33}, x_1, e_{13})] \\
\quad\quad\quad\quad\quad [(\text{*d})\ Coref(x_3, e_{33}, b, e_{31})]
\end{array}
\qquad
\begin{array}{l}
revise'(e_{42}, t, p_4) \\
\quad t : teacher'(e_{41}, t) \\
\quad p_4 : paper'(e_{45}, p_4) \\
\quad\quad Poss'(e_{44}, x_4, p_4) \\
\quad\quad\quad x_4 : he'(e_{43}, x_4) \\
\quad\quad\quad\quad [Coref(x_4, e_{43}, x_2, e_{23})\ (\text{*e})] \\
\quad\quad\quad\quad [Coref(x_4, e_{43}, x_3, e_{33})\ (\text{*f})] \\
\quad\quad\quad\quad [Coref(x_4, e_{43}, t, e_{41})\ (\text{*g})]
\end{array}
$$

Figure 3: Representations for Five Readings Case

(7) John realizes that he is a fool, but Bill does not, even though his wife does. (Dahl, 1972)

(8) John realizes that John is a fool, but Bill does not realize that Bill is a fool, even though Bill's wife realizes Bill is a fool.

Example (7) contains two ellipses. Reading (8) results from the second clause receiving a sloppy interpretation from the first, and the third clause receiving a strict interpretation from the second. An identity-of-relations analysis, however, predicts that this reading does not exist. Because the second clause will only have the sloppy derivation received from the first, the strict derivation that the third clause requires from the second will not be present.

However, in defending their identity-of-relations approach, Gawron and Peters (1990) note that a non-identity account predicts that sentence (9) has the (nonexistent) reading given in (10).

(9) John revised his paper before Bill did, but after the teacher did.

(10) John revised John's paper before Bill revised Bill's paper, but after the teacher revised John's paper.

In this case, the first clause is the antecedent for both ellipses. These two examples create a paradox; apparently neither type of analysis (nor any previous analyses we are aware of) can explain both.

Our analysis accounts for both examples through a mutually-constraining interaction of parallelisms. Example (7) is fairly straightforward, so we focus on example (9). Let us refer to the clauses as clauses 1, 2, and 3. Because clauses 2 and 3 are VP-elliptical, we must establish a parallelism between each of them and clause 1. In addition, the contrast relation signalled by "but" is justified by the contrasting predicates "before" and "after", provided their corresponding pairs of arguments are similar. Their first arguments are similar since they are identical—clause 1. Then we also have to establish the similarity of their second arguments—clause 2 and clause 3. Thus, three mutually constraining parallelisms must be established: 1 – 2, 1 – 3, and 2 – 3.

In Figure 4, cases *a and *b arise from the coreference and similarity options when establishing the parallelism between clauses 1 and 2, and cases *c and *d from the parallelism between clauses 1 and 3. However, because parallelism is also required between clauses 2 and 3, we cannot choose these options freely. If we choose case *a, then we must choose case *c, giving us the JJJ reading. If we choose case *b, then we must choose case *d, giving us the JBT reading. Because of the mutual constraints of the three parallelisms, no other readings are possible. This is exactly the right result.

Prüst (1992) essentially follows Sag's (1976) treatment of strict and sloppy readings, which, like other identity-of-relations analyses, will not generate the reading of the cascaded ellipsis sentence (7) shown in (8). While the approach will correctly predict the lack of reading (10) for sentence (9), it does so for the wrong reason. Whereas ellipsis resolution does not permit such readings in any circumstance in his account, we claim that the lack of such readings for sentence (9) is due to constraints imposed by multiple parallelisms, and not because of the correctness of identity-of-relations analyses.

Asher's (1993) analysis falls into the non-identity class of analyses, and therefore makes the correct predictions for sentence (7). While he does not discuss the contrast between this case and sentence (9), we do not see any reason why his framework could not accommodate our solution.

6 Other Examples

Missing Readings with Multiple Pronouns Dahl (1974) noticed that sentence (11) has only three readings instead of the four one might expect. The reading *Bill said that John revised Bill's paper* is missing.

(11) John said that he revised his paper, and Bill did too.

$$
\begin{array}{ll}
before(e_{12}, e_{22}) & \\
\quad e_{12} : revise'(e_{12}, j, p_1) & e_{22} : revise'(e_{22}, b, p_2) \\
\quad\quad j : John'(e_{11}, j) & \quad b : Bill'(e_{21}, b) \\
\quad\quad p_1 : paper'(e_{15}, p_1) & \quad p_2 : paper'(e_{25}, p_2) \\
\quad\quad\quad Poss'(e_{14}, x_1, p_1) & \quad\quad Poss'(e_{24}, x_2, p_2) \\
\quad\quad\quad\quad x_1 : he'(e_{13}, x_1) & \quad\quad\quad x_2 : he'(e_{23}, x_2) \\
\quad\quad\quad\quad\quad Coref(x_1, e_{13}, j, e_{11}) & \quad\quad\quad\quad [Coref(x_2, e_{23}, x_1, e_{13}) \text{ (*a)}] \\
 & \quad\quad\quad\quad [Coref(x_2, e_{23}, b, e_{21}) \text{ (*b)}]
\end{array}
$$

$$
\begin{array}{l}
after(e_{12}, e_{32}) \\
\quad e_{32} : revise'(e_{32}, t, p_3) \\
\quad\quad t : teacher'(e_{31}, t) \\
\quad\quad p_3 : paper'(e_{35}, p_3) \\
\quad\quad\quad Poss'(e_{34}, x_3, p_3) \\
\quad\quad\quad\quad x_3 : he'(e_{33}, x_3) \\
\quad\quad\quad\quad\quad [Coref(x_3, e_{33}, x_1, e_{13}) \text{ (*c)}] \\
\quad\quad\quad\quad\quad [Coref(x_3, e_{33}, t, e_{31}) \text{ (*d)}]
\end{array}
$$

Figure 4: Representations for the Source-of-Ellipsis Paradox

In contrast, the similar sentence given in (12) appears to have all four readings.

(12) John said that his teacher revised his paper, and Bill did too.

The readings derived by our analysis depend on the *Coref* relations that hold between the coreferring noun phrases in the antecedent clauses. For sentence (11), the correct readings result if *his* is linked to *he* and *he* to *John*; for sentence (12), the correct readings result if both pronouns are linked to *John*. Other cases in the literature indicate that the situation is more complicated than might initially be evident. Handling these cases requires an account of how such dependencies are established, which we discuss in Hobbs and Kehler (forthcoming).

Extended Parallelism In some cases, the elements involved in a sloppy reading may not be contained in the minimal clause containing the ellipsis.

(13) John told a man that Mary likes him, and Bill told a boy that Susan does.[7]

Although the antecedent clause for "Susan does" is "Mary likes him", there is a sloppy reading in which "Bill told a boy that Susan likes Bill". This fact is problematic for accounts of VP ellipsis that operate only within the minimal clauses. These readings are predicted by our account, as John and Bill are parallel in the main clauses.

Lazy Pronouns "Lazy pronouns" can be accounted for similarly. In

(14) The man who gives his paycheck to his wife is wiser than the man who gives it to his mistress. (Karttunen, 1969)

the pronoun *it* does not refer to the first man's paycheck but the second's.

In text, *it* normally requires an explicit, coreferring antecedent. However, the parallelism between the clauses licenses a sloppy reading via the similarity option. The real world fact that to give something to someone, you first must have it, leads to a strong preference for the sloppy reading.

It is necessary to have parallelism in order to license the lazy pronoun reading. If we eliminate the possibility of parallelism, as in

(15) John revised his paper, and then Bill handed it in.

the lazy pronoun reading is not available, even though the have-before-give constraint is not satisfied. To interpret this sentence, we are more likely to assume an unmentioned transfer event between the two explicit events.

Sloppy Readings with Events Sentence (16) has a "sloppy" reading in which the second main clause means "I will kiss you even if you don't want me to kiss you."

(16) I will help you if you want me to, but I will kiss you even if you don't.[8]

Deriving this reading requires a *Coref* relation between the elided event and its antecedent in the first main clause, which is obtained when our algorithm bails out in event coreference (see footnote

[7]This example is due to Prüst (1992), whose approach successfully handles this example.

[8]Mark Gawron, p.c., attributed to Carl Pollard.

3). Then in expanding the VP ellipsis in the second main clause, taking the similarity option for the event generates the desired reading.

Inferentially-Determined Antecedents Webber (1978) provides several examples in which the antecedent of an ellipsis is derived inferentially:

(17) Mary wants to go to Spain and Fred wants to go to Peru, but because of limited resources, only one of them can.

Our account of parallelism applies twice in handling this example, once in creating a complex antecedent from recognizing the parallelism between the first two clauses, and again in resolving the ellipsis against this antecedent. Hobbs and Kehler (forthcoming) describe the analysis of this case as well as others involving quantification.

7 Summary

We have given a general account of parallelism in discourse and applied it to the special case of resolving possible readings for instances of VP ellipsis. In doing so, we showed how a variety of examples that have been problematic for previous approaches are accounted for in a natural and straightforward fashion. Furthermore, the generality of the approach makes it directly applicable to a variety of other types of ellipsis and reference in natural language.

Acknowledgements

The authors thank Mark Gawron, David Israel, and three anonymous reviewers for helpful comments. This research was supported by National Science Foundation/Advanced Research Projects Agency Grant IRI-9314961.

References

Asher, Nicholas. 1993. *Reference to Abstract Objects in Discourse.* SLAP 50, Dordrecht, Kluwer.

Crouch, Richard. 1995. Ellipsis and quantification: A substitutional approach. In *Proceedings of EACL-95*, pages 229–236, Dublin, Ireland, March.

Dahl, Östen. 1972. On so-called "sloppy" identity. *Gothenburg Papers in Theoretical Linguistics*, 11. University of Göteborg.

Dahl, Östen. 1974. How to open a sentence: Abstraction in natural language. *In Logical Grammar Reports, No. 12.* University of Göteborg.

Dalrymple, Mary, Stuart M. Shieber, and Fernando Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14:399–452.

Fiengo, Robert and Robert May. 1994. *Indices and Identity.* MIT Press, Cambridge, MA.

Gawron, Mark and Stanley Peters. 1990. *Anaphora and Quantification in Situation Semantics.* CSLI/University of Chicago Press, Stanford University. CSLI Lecture Notes, Number 19.

Hardt, Daniel. 1992. VP ellipsis and contextual interpretation. In *Proceedings COLING-92*, Nantes.

Hobbs, Jerry R.. 1979. Coherence and coreference. *Cognitive Science*, 3:67–90.

Hobbs, Jerry R. 1985. On the coherence and structure of discourse. Technical Report CSLI-85-37, Center for the Study of Language and Information, Stanford University, October.

Hobbs, Jerry R. and Andrew Kehler. Forthcoming. A general theory of parallelism and the special case of VP ellipsis. Technical report, SRI International.

Hobbs, Jerry R., Mark E. Stickel, Douglas E. Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63:69–142.

Karttunen, Lauri. 1969. Pronouns and variables. In *Papers from the Fifth Regional Meeting of the Chicago Linguistics Society.*

Kehler, Andrew. 1993. A discourse copying algorithm for ellipsis and anaphora resolution. In *Proceedings of EACL-93*, pages 203–212, Utrecht, the Netherlands, April.

Kehler, Andrew. 1995. *Interpreting Cohesive Forms in the Context of Discourse Inference.* Ph.D. thesis, Harvard University.

Lappin, Shalom and Michael McCord. 1990. Anaphora resolution in slot grammar. *Computational Linguistics*, 16:197–212.

Prüst, Hub. 1992. *On Discourse Structuring, VP Anaphora, and Gapping.* Ph.D. thesis, University of Amsterdam.

Sag, Ivan. 1976. *Deletion and Logical Form.* Ph.D. thesis, MIT.

Scha, Remko and Livia Polanyi. 1988. An augmented context free grammar for discourse. In *Proceedings of COLING-88*, pages 573–577, Budapest, August.

Webber, Bonnie Lynn. 1978. *A Formal Approach to Discourse Anaphora.* Ph.D. thesis, Harvard University.

Williams, Edwin. 1977. Discourse and logical form. *Linguistic Inquiry*, 8(1).